Numerical Approximations for Stochastic Differential Games: The Ergodic Case

Harold J. Kushner*
Applied Mathematics Department
Lefschetz Center for Dynamical Systems
Brown University, Providence, RI 02912
hjk@dam.brown.edu

December, 2001

Abstract

The Markov chain approximation method is a widely used, relatively easy to use, and efficient family of methods for the bulk of stochastic control problems in continuous time, for reflected-jump-diffusion type models. It has been shown to converge under broad conditions, and there are good algorithms for solving the numerical problems, if the dimension is not too high. We consider a class of stochastic differential games with a reflected diffusion system model and ergodic cost criterion and where the controls for the two players are separated in the dynamics and cost function. It is shown that the value of the game exists and that the numerical method converges to this value as the discretization parameter goes to zero. The actual numerical method solves a stochastic game for a finite state Markov chain and ergodic cost criterion. The essential conditions are nondegeneracy and that a weak local consistency condition hold "almost everywhere" for the numerical approximations, just as for the control problem.

1 Introduction

The Markov chain approximation method of [19, 20, 22] is a widely used method for the numerical solution of virtually all of the standard forms of stochastic control problems with reflected-jump-diffusion models. It is robust and can be shown to converge under very broad conditions. Extensions to approximations for two-person differential games with discounted, finite time, stopping time, and pursuit-evasion games were given in [18] for reflected diffusion models where the controls for the two players are separated in the dynamics and cost rate functions. In this paper, the basic ideas will be extended to two-player

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **DEC 2001**	2. REPORT TYPE	3. DATES COVERED **00-12-2001 to 00-12-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Numerical Approximations for Stochastic Differential Games: The Ergodic Case	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Brown University,Division of Applied Mathematics,182 George Street,Providence,RI,02912**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **26**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

stochastic dynamic games with the same systems model, but where the cost function is ergodic. Such ergodic and "separated" models occur, for example, in risk-sensitive and robust control [2, 3, 7, 15]. In fact, the game formulation of risk sensitive control problems for queues in heavy traffic was our original motivation.

When the robust control is for controlled queues in heavy traffic, then the state is confined to some convex polyhedron by boundary reflection [21]. In many other applications, the state of the physical problem is confined to a bounded set. One example is the heavy traffic limit of controlled queueing networks with finite buffers [1, 21] or robust control of such systems as in [2, 3], where the set is a hyperectangle. Then robust control would lead to a game problem with a hyperrectangular state space. If the system state is not a priori confined to a bounded set, then for numerical purposes it is commonly necessary to bound the state space artificially by adding a reflecting boundary and then experimenting with the bounds. Our systems model is confined to a state space G that is a convex polyhedron, and it is confined by a "reflection" on the boundary. More generally, the boundaries could be determined by a set of smooth curved surfaces as in [22], but we restrict attention to the polyhedral case, since that is the most common and it avoids minor details which can be distracting.

There are many results for various forms of the game problem; e.g., [4, 5, 6, 24, 28, 29]. But there seems to be nothing available concerned with the ergodic problem for the reflected diffusion model. We will use purely probabilistic methods of proof. Such methods have the advantage of providing intuition concerning numerical approximations, they cover many of problem formulations to date, and they converge under quite general conditions. The essential conditions are *weak-sense* existence and uniqueness of the solution to the controlled equations, "almost everywhere" continuity of the dynamical and cost rate terms, and a natural "local consistency" condition: The local consistency and continuity need hold only almost everywhere with respect to the measure of the basic model, hence discontinuities in the dynamics and cost function can be treated under appropriate conditions (see, in particular the treatment of discontinuities and complex variational problems with singularities and Theorems 4.6 and 7.1 in [22]). Furthermore, the numerical approximations are represented as processes which are close to the original, which gives additional intuitive and practical meaning to the method.

The methods to be used for the ergodic cost function are quite different than those used in [18]. They share the foundation in the theory of weak convergence [9, 13]. But they depend heavily on the approximations to the ergodic cost control problem as developed in [21, Chapter 4]. The development of the paper has been structured to take advantage of the results in [21, 22], wherever possible. To facilitate the development, Subsection 2.2 summarizes the results from [21] which will be needed here, with an occasional change of notation to suit that used here.

Subsection 2.1 defines the basic systems model, where the control is introduced via the Girsanov transformation [17]. The dynamical model is the

2

reflected stochastic differential equation (2.4), also called the *Skorohod problem* [12, 21, 22]. The conditions on the boundary of the state space are (A2.1)– (A2.2). Condition (A2.1) covers the great majority of cases of current interest, including those that arise from queueing and communications networks. The condition is obvious when the state space is a hyperrectangle with reflection directions being the interior normals. The strategies of the players are as follows. Player 1 wishes to minimize and player 2 to maximize. For the infsup problem (the upper value), at the start of the game (i.e., at $t = 0$) player 1 selects a control. This can be either a pure (and time independent) feedback control or a relaxed feedback control (see Subsection 2.1 for the definition). The selected control will be used at all $t \geq 0$. Then player 2 selects its strategy. This can be either a relaxed feedback or a classical relaxed control. Whatever it is, once selected, it cannot be changed.

The situation is analogous if player 2 selects first. Since the controls for the player who chooses first are time independent feedback and these are selected and fixed at the start of the game, and only the player choosing last can use time dependent controls, complications due to the notions of strategy in the time dependent case (e.g., concerning the definition of the value either via a limit of a discrete time game, or via the Elliott-Kalton definition) do not arise. In this sense the paper is simpler than [18]. On the other hand, the treatment of the ergodic cost criterion adds substantial new complications. Subsection 2.3 establishes the existence of the controls yielding the upper and lower values, using approximation methods from [21].

The Markov chain approximation numerical method is discussed in Subsection 3.1. The methods for getting the approximating chain and cost function are the same as in [22] for the pure control problem, since it is the process for arbitrary controls that is approximated. The natural *local consistency condition* is stated. The proof of convergence of the numerical method is in Subsection 3.2 and depends on the fact that the original game has a value. The numerical approximations are games for Markov chains. They might or might not have a value, depending on the form of the approximation. But, it it seen that the upper and lower values converge to the value of the original game as the approximation parameter goes to its limit. Finally, the proof that the original game has a value is given in Section 4.

2 The Dynamical Model and Background Results

2.1 Assumptions and the Dynamical Model

Assumptions. The first assumptions define the state space G.

A2.1. *The state space G is the intersection of a finite number of closed half spaces in Euclidean r-space $\mathbb{R}^r$, and is the closure of its interior (i.e., it is a*

*closed convex polyhedron with an interior and planar sides). Let ∂G_i, $i = 1, \ldots,$
denote the faces of G, and n_i the interior normal to ∂G_i. Interior to ∂G_i, the
reflection direction is denoted by the unit vector d_i, and $\langle d_i, n_i \rangle > 0$ for each i.
The possible reflection directions at points on the intersections of the ∂G_i are
in the convex hull of the directions on the adjoining faces. Let $d(x)$ denote the
set of reflection directions at the point $x \in \partial G$, whether it is a singleton or not.
No more than r constraints are active at any boundary point.*

A2.2. *For each $x \in \partial G$, define the index set $I(x) = \{i : x \in \partial G_i\}$. Suppose
that $x \in \partial G$ lies in the intersection of more than one boundary; that is, $I(x)$ has
the form $I(x) = \{i_1, \ldots, i_k\}$ for some $k > 1$. Let $N(x)$ denote the convex hull
of the interior normals $n_{i_1}, \ldots, n_{i_k}$ to $\partial G_{i_1}, \ldots, \partial G_{i_k}$, respectively, at x. Then,
there is some vector $v \in N(x)$ such that $\gamma' v > 0$ for all $\gamma \in d(x)$.*

   *There is a neighborhood $N(\partial G)$ and an extension of $d(\cdot)$ to $\overline{N(\partial G)}$ that is
upper semicontinuous in the following sense: For each $\epsilon > 0$, there is $\rho > 0$ that
goes to zero as $\epsilon \to 0$ and such that if $x \in N(\partial G) - \partial G$ and $distance(x, \partial G) \leq \rho$,
then $d(x)$ is in the convex hull of the directions $\{d(v); v \in \partial G, distance(x, v) \leq
\epsilon\}$.*

   Let $\alpha = (\alpha_1, \alpha_2), \alpha_1 \in U_1, \alpha_2 \in U_2$, denote the canonical control value, with
α_i the canonical value for player i.

A2.3. *The $U_i, i = 1, 2$, are compact sets in some Euclidean space. The $(r \times r)$
matrix-valued function $\sigma(\cdot)$ on G is Hölder continuous, with $\sigma^{-1}(x)$ bounded,
and the $\mathbb{R}^r$-valued functions $b_i(\cdot)$ on $G \times U_i$ are continuous.*

   The uncontrolled model is the solution to the Skorohod problem

$$dx(t) = \sigma(x(t))dw(t) + dz(t), \; x(t) \in G. \tag{2.1}$$

By a solution to (2.1) we mean the following. Let Ω denote the path space of
$(x(\cdot), z(\cdot), w(\cdot))$, and let $\{\mathcal{F}_t, t < \infty)\}$ denote the filtration on the space. The
$x(\cdot)$ and $z(\cdot)$ are $\mathbb{R}^r$-valued, continuous and $\mathcal{F}_t$-adapted, and $w(\cdot)$ is an $\mathcal{F}_t$-
standard $\mathbb{R}^r$-valued Wiener process. The $z(\cdot)$ is the reflection process. Let Ω_T
denote the restriction of Ω to functions defined on $[0, T]$. Define $\mathcal{F} = \lim_t \mathcal{F}_t$
and let P_x denote the measure when the initial condition is $x(0) = x$, with E_x
the associated expectation. Let $P_{x,T}(\cdot)$ denote the probability measure, when
we confine our interest to paths on the finite interval $[0, T]$.

   The controlled system will be defined via the Girsanov transformation, start-
ing with (2.1). For a detailed discussion of the Skorohod problem and the as-
sumptions (A2.1) and (A2.2), see [21, Chapter 3]. See also the brief comment
below (A2.4). We will also need the following condition.

A2.4. *There is a unique weak sense solution to (2.1) for each initial condition.*

Comments on (A2.1) and (A2.2). One can always construct the extension
in (A2.2). To see that (A2.1) is natural in application note the following. If the

state space is being bounded for purely numerical reasons, then the reflections are introduced only to give a compact set G, which should be large enough so that the effects on the solution in the region of main interest are small. A common choice is a hyperrectangle with normal reflection directions, in which case the right side of (2.1) is zero. Next, consider a queueing network model in the heavy traffic limit [16, 21, 27] where the state space is the nonnegative orthant, and the probability that an output of the ith processor goes to the jth processor is q_{ij}. If the spectral radius of the routing matrix $Q = \{q_{ij}; i, j\}$ is less than unity, then all customers will eventually leave the system. The model is a special case of (2.4) with $z(t) = [I - Q']y(t)$, where $y_i(\cdot)$ is nondecreasing, continuous, and can increase only at t where $x_i(t) = 0$. The condition (A2.1) implies (see [12, 21]) the so-called "completely-S" condition [16, 21, 26] which is used to ensure that $z(\cdot)$ has bounded variation w.p.1.

Classes of controls. A: Relaxed controls $r_i(\cdot)$. Suppose that for some filtration $\{\mathcal{F}_t, t < \infty\}$ and standard vector-valued $\mathcal{F}_t$-Wiener process $w(\cdot)$, each $r_i(\cdot), i = 1, 2$, is a measure on the Borel sets of $U_i \times [0, \infty)$ such that $r_i(U_i \times [0, t]) = t$ and $r_i(A \times [0, t])$ is $\mathcal{F}_t$-measurable for each Borel set $A \subset U_i$. Then $r_i(\cdot)$ is said to be an *admissible relaxed control* for player i, with respect to $w(\cdot)$. If the Wiener process and filtration have been given or are obvious or unimportant, then we simply say that $r_i(\cdot)$ is an admissible relaxed control for player i [14, 21, 22]. For Borel sets $A \subset U_i$, we will write $r_i(A \times [0, t]) = r_i(A, t)$.

For almost all (ω, t) and each Borel $A \subset U_i$, one can define the derivative

$$r_{i,t}(A) = \lim_{\delta \to 0} \frac{r_i(A, t) - r_i(A, t - \delta)}{\delta}.$$

Without loss of generality, we can suppose that the limit exists for each (ω, t). Then for all (ω, t), $r_{i,t}(\cdot)$ is a probability measure on the Borel sets of U_i and for any bounded Borel set B in $U_i \times [0, \infty)$,

$$r_i(B) = \int_0^\infty \int_{U_i} I_{\{(\alpha_i, t) \in B\}} r_{i,t}(d\alpha_i) dt.$$

An ordinary control $u_i(\cdot)$ can be represented in terms of the relaxed control $r_i(\cdot)$, defined by its derivative $r_{i,t}(A) = I_A(u_i(t))$, where $I_A(u_i)$ is unity if $u_i \in A$ and is zero otherwise. The weak topology [22] will be used on the space of admissible relaxed controls. Relaxed controls are commonly used in control theory to prove existence theorems, since any sequence of relaxed controls has a convergent subsequence.

B: Relaxed feedback control $m_i(\cdot)$ [10, 21]. Suppose that $m_i(x, \cdot), i = 1, 2$, is a probability measure on the Borel sets of U_i for each $x \in G$ and that $m_i(\cdot, A)$ is Borel measurable for each Borel set $A \subset U_i$. Then we say that $m_i(\cdot)$ is a relaxed feedback control. Define $U = U_1 \times U_2$. For relaxed feedback controls $m_i(\cdot)$, define $m(\cdot)$ by $m(x, d\alpha) = m_1(x, d\alpha_1)m_2(x, d\alpha_2)$. Then $m(\cdot)$ is also a relaxed

feedback control, but with control value space U. All $m(\cdot)$ will be of this product form for some relaxed feedback controls $m_i(\cdot), i = 1, 2$. If $x(\cdot)$ is a solution to (2.4), and $m(\cdot)$ a relaxed feedback control, then $m(\cdot)$ can be represented by a relaxed control $r(\cdot)$ with derivative $r_t(d\alpha) = r_{1,t}(d\alpha_1)r_{2,t}(d\alpha_2) = m(x(t), d\alpha)$.

The control for the player that chooses its control first will always be a relaxed feedback control, but that for the player who chooses its control last might be either a relaxed feedback control or a relaxed control which is not representable in relaxed feedback form.

Defining the controlled dynamical system via the Girsanov transformation: Relaxed feedback controls.

The controlled model will be defined via the Girsanov transformation [17]. Some of the well known details will be described, since the equations will be needed for the approximations. This will be done first for the relaxed feedback controls. Let $m_i(\cdot), i = 1, 2$, be relaxed feedback controls and define $m(x, d\alpha) = m_1(x, d\alpha_1)m_2(x, d\alpha_2)$. Define

$$b_{i,m_i}(x) = \int_{U_i} b_i(x, \alpha_i)m_i(x, d\alpha_i), \quad b(x, \alpha) = b_1(x, \alpha_1) + b_2(x, \alpha_2),$$

and set $b_m(x) = \int_U b(x, \alpha)m(x, d\alpha) = b_{1,m_1}(x) + b_{2,m_2}(x)$. For $T > 0$ and relaxed feedback control $m(\cdot)$, define

$$\zeta(T, m) = \int_0^T \left[\sigma^{-1}(x(s))b_m(x(s))\right]' dw(s) - \frac{1}{2}\int_0^T \left|\sigma^{-1}(x(s))b_m(x(s))\right|^2 ds,$$

and set

$$R(T, m) = e^{\zeta(T, m)}.$$

For each $(x, T, m(\cdot))$, define the measure $P_{x,T}^m$ on $(\Omega_T, \mathcal{F}_T)$ via the Radon–Nikodym derivative $R(T, m)$:

$$dP_{x,T}^m = R(T, m)dP_{x,T}. \tag{2.2}$$

For each $(x, m(\cdot))$, the family $P_{x,T}^m$ of measures, indexed by T, is consistent and can be extended uniquely to a measure P_x^m on $(\Omega, \mathcal{F})$ that is consistent with the $P_{x,T}^m$. When there is no control (i.e., where the system is (2.1)), we omit the superscript m. The process $w_m(\cdot)$ defined by

$$dw_m(t) = dw(t) - \left[\sigma^{-1}(x(s))b_m(x(s))\right] dt \tag{2.3}$$

is an $\mathcal{F}_t$-standard Wiener process on $(\Omega, P_x^m, \mathcal{F})$ [17]. Now, rewrite the uncontrolled model (2.1) as

$$dx(t) = b_m(x(t))dt + \sigma(x(t))dw_m(t) + dz(t). \tag{2.4}$$

Under the measures $\{P_x^m, x \in G\}$, (2.4) is a Markov process and we use $P^m(x, t, \cdot)$ for its transition function. Use $P(x, t, \cdot)$ for the transition function of the uncontrolled process (2.1). Strictly speaking, the process $w_m(\cdot)$ should be indexed also by the initial condition $x = x(0)$, but we omit it for notational simplicity.

6

The controlled dynamical system with relaxed controls. Let $r_i(\cdot)$ be a relaxed control for player i, with derivative $r_{i,t}(\cdot)$, and define $b_{i,r_i}(x,t) = \int_{U_i} b_i(x,\alpha) r_{i,t}(d\alpha_i)$. We will also have occasion to use relaxed (and not necessarily relaxed feedback) controls for one of the players. For specificity at this point, suppose that a relaxed control is used for player 1 and a relaxed feedback control is used for player 2. Write $b_{r_1,m_2}(x,t) = b_{1,r_1}(x,t) + b_{2,m_2}(x)$, define $\xi(T,r_1,m_2), P_{x,T}^{r_1,m_2}$, $P_x^{r_1,m_2}$, and $w_{r_1,m_2}(\cdot)$ analogously to what was done for the pure relaxed feedback control case, and rewrite the controlled equation as

$$dx(t) = b_{1,r_1}(x(t),t)dt + b_{2,m_2}(x)dt + \sigma(x(t))dw_{r_1,m_2}(t) + dz(t). \qquad (2.5)$$

The measures $P_x^{r_1,m_2}$ are used with (2.5). The development is analogous if player 1 uses the relaxed feedback control and player 2 the relaxed control.

Representation of the reflection process $z(\cdot)$. For either the model (2.4) or (2.5), the process $z(\cdot)$ can be represented as

$$z(t) = \sum_i y_i(t)d_i, \qquad (2.6)$$

where $y_i(\cdot)$ is nondecreasing, right continuous, increases only at t where $x(t)$ is on the i-th face of G and satisfies $y_i(0) = 0$. Under (A2.1), (A2.2), and (A2.4), the representation (2.6) is unique with probability one [21, Theorem 3.6, Chapter 4]. Let M_ϵ denote an ϵ-neighborhood of the boundary set where more than one constraint is active. Then, the same theorem implies that, for $t > 0$, $\sup_{x,m} E_x^m |y(t)| I_{\{x(t) \in M_\epsilon\}} \to 0$ as $\epsilon \to 0$.

2.2 Background Results and the Cost Function

The development depends heavily on approximation, continuity, and limit results from [21, Chapter 4] for the control problem. The results carry over to the game problem, since they are concerned with arbitrary relaxed feedback and relaxed controls. To facilitate our development, several key results from [21] will be stated, in the notation of this paper.

Illustration of the use of the Girsanov transformation: Mutual absolute continuity of the transition functions. The following theorem is [21, Theorem 3.1, Chapter 4]. We will outline the proof by copying some of the details from the reference, since similar "Girsanov transformation" methods underlie many of the results, there are some slight differences worth noting, and it gives a feeling for the approach. Unless otherwise noted, "almost all" refers to Lebesgue measure. The symbol $\Rightarrow$ denotes weak convergence.

Theorem 2.1. *Assume* (A2.1)–(A2.4). *Let* $m^n(y,\cdot) \Rightarrow m(y,\cdot)$ *for almost all* $y \in G$, *where* $m(\cdot)$ *and* $m^n(\cdot)$ *are relaxed feedback controls. Then for any* $0 < t_0 < t_1 < \infty$ *and bounded and measurable real-valued function* $f(\cdot)$,

$$\int f(y)P^{m^n}(x,t,dy) \to \int f(y)P^m(x,t,dy) \qquad (2.7)$$

uniformly for $(x,t) \in G \times [t_0, t_1]$. For any $t > 0$, $P^m(x,t,\cdot)$ is absolutely continuous with respect to Lebesgue measure, uniformly in $m(\cdot)$ and in $(x,t) \in G \times [t_0, t_1]$. For each relaxed feedback control $m(\cdot)$, the process defined by (2.4) is a strong Feller process and it has a unique weak-sense solution for each initial condition x.

Proof. We concentrate on the uniformity in x of the convergence (2.7). First note that, by the weak convergence and the product form of $m^n(\cdot)$, the limit $m(\cdot)$ can always be represented as $m(x, d\alpha) = m_1(x, d\alpha_1)m_2(x, d\alpha_2)$ for some relaxed feedback controls $m_i(\cdot), i = 1, 2$, for almost all x. The expression (2.7) can be written equivalently as

$$E_x f(x(t))R(t, m^n) - E_x f(x(t))R(t, m) \to 0. \tag{2.8}$$

For notational simplicity, let $\sigma(x) = I$, the identity. We will use the inequalities:

$$\left| e^a - e^b \right| \le |a - b| \left[e^a + e^b \right], \tag{2.9a}$$

$$E_x \left| \int_0^t b'_m(x(s))dw(s) - \int_0^t b'_{m^n}(x(s))dw(s) \right|^2 \tag{2.9b}$$
$$\le E_x \int_0^t |b_m(x(s)) - b_{m^n}(x(s))|^2 ds.$$

By the continuity and boundedness of $b(\cdot)$ and the weak convergence of the $m^n(y, \cdot)$ for almost all $y \in G$, we have

$$b_{m^n}(y) = \int_U b(y, \alpha)m^n(y, d\alpha) \to b_m(y) = \int_U b(y, \alpha)m(y, d\alpha)$$

for almost all y. Define

$$\tilde{b}_n(y) = |b_m(y) - b_{m^n}(y)|^2.$$

Let $t \in [t_0, t_1]$, where $0 < t_0 < t_1 < \infty$. By Egoroff's theorem [11, Theorem 12, page 149], for each $\epsilon > 0$, there is a measurable set A_ϵ with $l(A_\epsilon) \le \epsilon$ such that $\tilde{b}_n(y) \to 0$ uniformly in $y \notin A_\epsilon$. Furthermore, $P(x, t, \cdot)$ is absolutely continuous with respect to Lebesgue measure for each x and $t > 0$ (and uniformly in $(x, t) \in G \times [t_0, t_1]$ for any $0 < t_0 < t_1 < \infty$). These facts imply that

$$\int_0^t E_x \tilde{b}_n(x(s))ds \to 0,$$

uniformly in $x \in G$. The last expression, together with the inequalities (2.9), implies (2.8) uniformly in $x \in G$.

Additional background results. We will also need the results of Theorems 2.2 to 2.8, most of which are either taken from [21] or are minor adaptations of such results. Where an elaboration on a proof in [21] would be useful, additional

comments will be made. Although the reference does not deal with games, the fact that the product $m(x, d\alpha) = m_1(x, d\alpha_1)m_2(x, d\alpha_2)$ is a relaxed feedback control allows the results to be carried over.

Theorem 2.2. (From [21, Theorems 3.1–3.3, Chapter 4].) *Assume* (A2.1)–(A2.4). *The process $x(\cdot)$ defined by (2.4) has a unique invariant measure $\mu_m(\cdot)$ for each relaxed feedback control $m(x, d\alpha) = m_1(x, d\alpha_1)m_2(x, d\alpha_2)$. Furthermore the transition function $P^m(x, t, \cdot)$ is mutually absolutely continuous with respect to Lebesgue measure, uniformly in $m(\cdot), x \in G$, and $t \in [t_0, t_1]$ for any $0 < t_0 < t_1 < \infty$.*

A smoothed control. Extend the definition of the relaxed feedback control $m_i(y, \cdot)$ so that it is defined as a relaxed feedback control for all $y \in \mathbb{R}^r$. For example, let it be concentrated on some fixed number in U for $y \notin G$. For small $\epsilon > 0$ and $x \in G$, define the smoothed control

$$m_{i,\epsilon}(x, \cdot) = \frac{1}{(2\pi\epsilon)^{r/2}} \int_{\mathbb{R}^r} e^{-|y-x|^2/2\epsilon} m_i(y, \cdot)dy, \quad x \in G.$$

Define $m_\epsilon(x, \cdot) = m_{1,\epsilon}(x, \cdot)m_{2,\epsilon}(x, \cdot)$.

Theorem 2.3. (This is [21, Theorem 3.4, Chapter 4].) *Assume* (A2.1)–(A2.4). *$m_\epsilon(\cdot)$ is a relaxed feedback control and $m_\epsilon(x, \cdot) \Rightarrow m(x, \cdot) = m_1(x, \cdot)m_2(x, \cdot)$ for almost all $x \in G$. The function $b_{m_2}(\cdot)$ is continuous for each ϵ, and $b_{m_2}(x) \to b_m(x)$ almost everywhere in G.*

Theorem 2.4. (From [21, Theorem 4.2, Chapter 4].) *Assume* (A2.1)–(A2.4). *Then $\mu_m(\cdot)$ is continuous in the control in that if $m^n(x, \cdot) \Rightarrow m(x, \cdot)$ for almost all $x \in G$, then for each Borel set $A \subset G$,*

$$\mu_{m^n}(A) \to \mu_m(A).$$

The cost function. We will need the following assumption.

A2.5. *The real-valued functions $k_i(\cdot)$ on $G \times U_i, i = 1, 2$, are continuous, and c is a vector with nonnegative components.*

Define $k(x, \alpha) = k_1(x, \alpha_1) + k_2(x, \alpha_2)$. For a relaxed feedback control $m(\cdot)$, define $k_m(x) = \int_U k(x, \alpha)m(x, d\alpha)$ and

$$\gamma_T(x, m) = \frac{1}{T}E_x^m \int_0^T k_m(x(s))ds + \frac{1}{T}E_x^m c'y(T).$$

For relaxed feedback controls, the cost function of interest in this paper is

$$\gamma(m) = \lim_T \gamma_T(x, m). \tag{2.10}$$

9

We omit the $x = x(0)$ from the argument of $\gamma(m)$, since it will not depend on the initial condition under our assumptions (see Theorem 2.5). If player i uses a relaxed control $r_i(\cdot)$, then define

$$k_{r_i}(x,t) = \int_{U_i} k_i(x, \alpha_i) r_{i,t}(d\alpha_i).$$

If player 1 selects its control first and uses a relaxed feedback control and player 2 selects its control last and uses a relaxed control, then define (the use of $\liminf$ is just a convention):

$$\gamma_T(x, m_1, r_2) = \frac{1}{T} E_x^{m_1, r_2} \int_0^T \left[k_{1,m_1}(x(s)) + k_{2,r_2}(x(s), s) \right] ds + \frac{1}{T} E_x^{m_1, r_2} c' y(T),$$

$$\gamma(x, m_1, r_2) = \liminf_T \gamma_T(x, m_1, r_2),$$

If player 2 selects its control first and uses a relaxed feedback control and player 1 uses a relaxed control, define (the use of $\limsup$ is just a convention):

$$\gamma(x, r_1, m_2) = \limsup_T \gamma_T(x, r_1, m_2).$$

Representation of the cost in terms of a stationary system. Let $m(\cdot)$ be a relaxed feedback control. The system (2.4) starts with an arbitrary initial condition that does not necessarily have the stationary distribution. It turns out that the limit (2.10) is the same as if the initial condition were distributed as $\mu_m(\cdot)$. This is the assertion of the next theorem.

Theorem 2.5. (This is [21, Theorem 4.1, Chapter 4].) *Assume* (A2.1)–(A2.5). *Let $m(\cdot)$ be a relaxed feedback control. Then the $E_x^m y_i(1)$ are continuous functions of x and*

$$\lim_T \gamma_T(x, m) = \gamma(m)$$
$$= \int k_m(x) \mu_m(dx) + \int E_x^m \left[c' y(1) \right] \mu_m(dx).$$

2.3 Existence of Optimal Controls for the Upper and Lower Values

Define the upper and lower values, resp., for the game (fb denotes relaxed feedback, and rel denotes relaxed controls)

$$\bar{\gamma}^+ = \inf_{\text{relaxed fb } m_1} \sup_{\text{rel controls } r_2} \gamma(m_1, r_2), \tag{2.11a}$$

$$\bar{\gamma}^- = \sup_{\text{relaxed fb } m_2} \inf_{\text{rel controls } r_1} \gamma(r_1, m_2). \tag{2.11b}$$

10

It is shown below that the use of relaxed controls for the player selecting last offers no advantage over feedback controls. In Section 4 it is shown that the game has a value in that $\bar{\gamma}^+ = \bar{\gamma}^- = \bar{\gamma}$. Then the numerical procedure converges to $\bar{\gamma}$ as the discretization level goes to zero (see Section 3).

The definition (2.11a) is interpreted to mean that player 2 supposes that player 1 has selected a relaxed feedback control for itself, which will be fixed throughout the game. [I.e., player 1 selects first.] Given this presumed choice of player 1, player 2 can select any relaxed or relaxed feedback control and will choose so as to maximize. This maximizing control will exist and will actually be of the relaxed feedback control form (implied by Theorem 2.8). It will depend on the presumed choice of player 1. Given this relationship, player 1 will select a minimizing control. By Theorem 2.8, it will exist and be of the relaxed feedback form. The interpretation of (2.11b) is analogous.

Theorem 2.6. (This is [21, Theorem 4.3, Chapter 4], adapted to the notation of the present case.) *Assume* (A2.1)–(A2.5). *For a sequence* $\{m^n(\cdot)\}$ *of relaxed feedback controls, let* $m^n(x, \cdot)$ *converge weakly to* $m(x, \cdot)$ *for almost all* $x \in G$. *Then* $\gamma(m^n) \to \gamma(m)$.

For fixed $m_1(\cdot)$, *maximize over* $m_2(\cdot)$, *and let* $\{m_2^n(\cdot)\}$ *be a maximizing sequence. Consider measures over the Borel sets of* $G \times U$ *which are defined by*

$$m^n(x, d\alpha)dx = m_1(x, d\alpha_1)m_2^n(x, d\alpha_2)dx \qquad (2.12)$$

and take a weakly convergent subsequence. The limit can be factored into the form

$$m_1(x, d\alpha_1)\tilde{m}_2(x, d\alpha_2)dx, \qquad (2.13)$$

where $\tilde{m}_2(\cdot)$ *is a relaxed feedback control for player 2. Since* $\tilde{m}_2(\cdot)$ *depends on* $m_1(\cdot)$, *write it as* $\tilde{m}_2(\cdot) = \overline{m}_2(\cdot; m_1)$. *Then, given* $m_1(\cdot)$, *the relaxed feedback control* $\overline{m}_2(\cdot; m_1)$ *is maximizing for player 2 in that*

$$\sup_{m_2} \gamma(m_1, m_2) = \gamma(m_1, \overline{m}_2(m_1))$$

The analogous result holds in the other direction, where player 2 chooses first.

Remark on the proof. First, note that owing to the product form any weak sense limit of the sequence defined in (2.12) must be of the form (2.13) where $\tilde{m}_1(\cdot)$ is a relaxed feedback control. The reference [21, Theorem 4.3, Chapter 4] is concerned with a minimization problem. Changing minimization to maximization and adapting the notation to our case where there are two controls and one is fixed, it shows that the limit $m_1(x, d\alpha_1)\tilde{m}_2(x, d\alpha_2)$ is maximizing, which is the assertion of the second paragraph of the theorem.

Relaxed controls for the player who chooses last. Suppose that with $m_1(\cdot)$ fixed, player 2 is allowed to use relaxed controls and not simply relaxed feedback controls. The following theorem says that the maximization over this

11

larger class will not yield a better result for player 2. The analog of the result for player 2 choosing first also holds.

Theorem 2.7. (This is [21, Theorem 6.1, Chapter 4], adapted to the notation of the present case.) *Assume* (A2.1)–(A2.5), *Fix* $m_1(\cdot)$ *and let* $\overline{m}_2(\cdot; m_1)$ *be an optimal relaxed feedback control and* $r_2(\cdot)$ *an arbitrary relaxed control for player 2. Then for each* $x \in G$,

$$\gamma(x, m_1, r_2) \leq \gamma(m_1, \overline{m}_2(m_1)).$$

Theorem 2.8. *Assume* (A2.1)–(A2.5). *Let player 1 go first. Then it has an optimal control, denoted by* $\overline{m}_1^+(\cdot)$. *The analogous result holds if player 2 chooses first, and its optimal control is denoted by* $\overline{m}^-(\cdot)$.

Remark on the proof. The proof is essentially a consequence of [21, Theorem 4.3, Chapter 4], just as Theorem 2.6 was. Let player 1 go first and let $\{m_1^n(\cdot)\}$ be a minimizing sequence of relaxed feedback controls. By Theorem 2.6, if player 1 uses $m_1^n(\cdot)$ then player 2 would use the (maximizing) relaxed feedback control $\overline{m}_2(\cdot; m_1^n)$. Following the method of the reference that was used to prove Theorem 2.6, take a weakly convergent subsequence of the sequence of measures on the Borel sets of $G \times U$ that is defined by $m_1^n(x, d\alpha_1)\overline{m}_2(x, d\alpha_2; m_1^n)dx$. and denote the limit by $\overline{m}_1^+(x, d\alpha_1)\tilde{m}_2(x, d\alpha_2)dx$. Any weak sense limit must have this form, where the $\overline{m}_1^+(\cdot)$ and $\tilde{m}_2(\cdot)$ are relaxed feedback controls. For notational simplicity, let n index the weakly convergent subsequence. Then, we must have $m_1^n(x, \cdot) \Rightarrow \overline{m}_1^+(x, \cdot)$ and $\overline{m}_2(x, \cdot; m_1^n) \Rightarrow \tilde{m}_2(x, \cdot)$ for almost all $x \in G$.

We need to show that $\overline{m}_1^+(\cdot)$ is optimal for player 1 if it chooses first, and that it can be supposed that $\tilde{m}_2(\cdot) = \overline{m}_2(\cdot; \overline{m}_1^+)$. Since $\{m_1^n(\cdot)\}$ is minimizing for player 1 when it chooses first, $\gamma(m_1^n, \overline{m}_2(m_1^n)) \to \bar{\gamma}^+$. Suppose that $\bar{\gamma}^+ < \sup_{m_2} \gamma(\overline{m}_1^+, m_2)$. Then there is $\hat{m}_2(\cdot)$ such that $\bar{\gamma}^+ < \gamma(\overline{m}_1^+, \hat{m}_2)$. Now, let player 2 use $\hat{m}_2(\cdot)$ instead of $\overline{m}_2(\cdot; m_1^n)$ for large n. Since the sequence defined by $m_1^n(x, d\alpha_1)\hat{m}_2(x, d\alpha_2)dx$ converges weakly to the measure defined by $\overline{m}_1^+(x, d\alpha_1)\hat{m}_2(x, d\alpha_2)dx$, Theorem 2.6 implies that $\gamma(m_1^n, \hat{m}_2) \to \gamma(\overline{m}_1^+, \hat{m}_2) > \bar{\gamma}^+$. This contradicts the fact that $\{m_1^n(\cdot)\}$ is minimizing, since it implies that there is $\epsilon > 0$ such that $\gamma(m_1^n, \hat{m}_2) \geq \bar{\gamma}^+ + \epsilon$ for large n. Thus $\overline{m}_1^+(\cdot)$ is optimal for player 1 if it chooses first. Since $\bar{\gamma}^+ = \gamma(\overline{m}_1^+, \tilde{m}_2)$, without loss of generality we can suppose that $\tilde{m}_2(\cdot) = \overline{m}_2(\cdot; \overline{m}_1^+)$.

Remark on smooth nearly optimal controls. In Section 4 we will need the fact that the optimal relaxed feedback controls for either player can be smoothed with little loss. In particular, suppose that player 1 chooses first, let $\epsilon > 0$, and replace $\overline{m}_1^+(\cdot)$ by the smoothed $\overline{m}_{1,\epsilon}^+(\cdot)$ as defined above Theorem 2.3. It is true that

$$\limsup_{\epsilon \to 0} \sup_{m_2} \gamma(\overline{m}_{1,\epsilon}^+, m_2) = \bar{\gamma}^+. \tag{2.14}$$

12

To prove (2.14), suppose that it does not hold in that there is $\delta > 0$ such that

$$\limsup_{\substack{\epsilon \to 0 \\ m_2}} \gamma(\overline{m}_{1,\epsilon}^+, m_2) \geq \bar{\gamma}^+ + \delta. \tag{2.15}$$

Then there are $m_{2,\epsilon}(\cdot)$ such that $\gamma(\overline{m}_{1,\epsilon}^+, m_{2,\epsilon}) \geq \bar{\gamma}^+ + \delta/2$ for all small $\epsilon > 0$. Let ϵ index a weakly convergent subsequence of $\overline{m}_{1,\epsilon}^+(x, d\alpha_1) m_{2,\epsilon}(x, d\alpha_2) dx$. The limit can be written as $\overline{m}_1^+(x, d\alpha_1) \tilde{m}_2(x, d\alpha_2) dx$ for some relaxed feedback control $\tilde{m}_2(\cdot)$. By Theorem 2.6, $\gamma(\overline{m}_{1,\epsilon}^+, m_{2,\epsilon}) \to \gamma(\overline{m}_1^+, \tilde{m}_2) \geq \bar{\gamma}^+ + \delta/2$, a contradiction to the optimality of $\overline{m}_1^+(\cdot)$ for player 1 if it chooses first. Obviously, there is an analog if player 2 chooses first.

3 Convergence of the Numerical Procedure

Discuss the connection.

3.1 The Markov Chain Approximation Method

The numerical method to be employed is the Markov chain approximation method of [19, 20, 22]. The approximating processes are the same. But the numerical problem to be solved is an ergodic cost problem for a Markov chain. The method approximates the system process (2.4) by a discrete parameter finite state controlled Markov chain that is "locally consistent" with (2.4). The cost function is also approximated and the game problem is then solved. Some basic facts from [22] concerning the procedure will now be stated. Let h denote the approximation parameter. Many methods for getting suitable approximating chains are in the references (e.g., see [22, Chapter 5]). The approximating chain and local consistency conditions are the same for the game problems of this paper. In the present case, where $\sigma(x)\sigma'(x)$ is uniformly positive definite, for each small fixed value of h the constructed chains can be selected to be ergodic for each control [22, Chapter 7]) and this will be assumed to be the case. In fact, the chains can be chosen such that for each small h, the rate of convergence of the transition functions to the invariant measure (as time goes to infinity) will be uniform in the control. See [22, Chapter 7] for a discussion of the setup and convergence for the pure control problem.

To construct the approximation, one first defines S_h, a discretization of $I\!\!R^r$. For example, S_h might be a regular $h-$grid. The precise requirements are quite weak and it is only the points in G and their immediate neighbors that are of interest. The state space for the chain is divided into two parts. The first part is $G_h = G \cap S_h$, on which the chain approximates the diffusion part of (2.4). If the chain tries to leave G_h, then it is returned immediately, consistently with the local reflection direction. Thus, define ∂G_h^+ to be the set of points not in G_h to which the chain might move in one step from some point in G_h. The set ∂G_h^+ is an approximation to the reflecting boundary. The use of ∂G_h^+ simplifies the analysis and allows us to get a reflection process $z^h(\cdot)$ that is analogous to $z(\cdot)$.

Local consistency on G_h. Let $u_n^h = (u_{1,n}^h, u_{2,n}^h)$ denote the controls used at step n for the approximating chain ξ_n^h. Let $E_{x,n}^{h,\alpha}$ (respectively, $\mathrm{covar}_{x,n}^{h,\alpha}$) denote the expectation (respectively, the covariance) given all of the data to step n, when $\xi_n^h = x, u_n^h = \alpha$. Then the chain satisfies the following consistency condition. There is $\Delta t^h(x,\alpha) = \Delta t^h \to 0$ (it does not depend on (x,α) for $x \in G$) such that

$$
E_{x,n}^{h,\alpha}\left[\xi_{n+1}^h - x\right] = b(x,\alpha)\Delta t^h + o(\Delta t^h),
$$
$$
\mathrm{covar}_{x,n}^{h,\alpha}\left[\xi_{n+1}^h - x\right] = a(x)\Delta t^h + o(\Delta t^h), \ a(x) = \sigma(x)\sigma'(x), \tag{3.1}
$$
$$
\|\xi_{n+1}^h - \xi_h^n\| \le K_1 h,
$$

for some real K_1. The $o(\Delta t^h)$ terms are uniform in (x,α). Let $P^h(x,y|\alpha_1,\alpha_2) = P^h(x,y|\alpha)$ denote the one-step transition probabilities. With the methods in [22], Δt^h is obtained automatically as a byproduct of getting the $P^h(x,y|\alpha)$, and it is used as an interpolation interval. More generally, Δt^h can depend on x,α. But for theoretical purposes for the ergodic cost problem, the problem is rescaled to get constant intervals. See the discussion in [22, Chapter 7]. By (3.1), in G the conditional mean first two moments of $\xi_{n+1}^h - \xi_n^h$ are close to those of the differences of the solution to (2.4).

The first two lines of (3.1) give the conditional moments for any fixed control values $\alpha = (\alpha_1, \alpha_2)$. Suppose that the control is chosen at random, depending only on the current state (i.e., it is randomized feedback). Let $m_i^h(x, d\alpha_i)$ denote the associated probability, conditioned on the past and on the current state value x, and define $m^h(x, d\alpha) = m_1^h(x, d\alpha_1) m_2^h(x, d\alpha_2)$. Then the transition probability is

$$
\int_U P^h(x, y|\alpha_1, \alpha_2) m_1^h(x, d\alpha_1) m_2^h(x, d\alpha_2).
$$

The first two lines of (3.1) are now replaced by

$$
E_{x,n}^{h,m^h}\left[\xi_{n+1}^h - x\right] = b_{m^h}(x)\Delta t^h + o(\Delta t^h),
$$
$$
\mathrm{covar}_{x,n}^{h,m^h}\left[\xi_{n+1}^h - x\right] = a(x)\Delta t^h + o(\Delta t^h), \ a(x) = \sigma(x)\sigma'(x). \tag{3.2}
$$

Thus, the forms are the same as if relaxed feedback controls were used. Although the actual sample paths would differ, the transition probabilities are the same for the randomized and the relaxed feedback forms.

Local consistency on ∂G_h^+. From points in ∂G_h^+, the transitions of the chain are such that they move to G_h, with the conditional mean direction being a reflection direction at x. More precisely,

$$
\lim_{h \to 0} \sup_{x \in \partial G_h^+} \mathrm{distance}(x, G_h) = 0, \tag{3.3}
$$

and there are $\theta_1 > 0$ and $\theta_2(h) \to 0$ as $h \to 0$ such that for all $x \in \partial G_h^+$,

$$
E_{x,n}^{h,\alpha}\left[\xi_{n+1}^h - x\right] \in \{a\gamma : \gamma \in d(x), \theta_2(h) \ge a \ge \theta_1 h\},
$$
$$
\Delta t^h(x,\alpha) = 0 \ \text{ for } x \in \partial G_h^+. \tag{3.4}
$$

14

The last line of (3.4) says that the reflection from states on ∂G_h^+ is instantaneous. Without loss of generality, we can suppose that the transition probabilities are continuous in the control variables for each x (see [22, Chapter 5] for typical methods of construction).

Continuous time interpolation. Only the discrete time chain ξ_n^h is needed for the numerical computations. But, for the proofs of convergence, the chain must be interpolated into a continuous time process which approximates $x(\cdot)$. The interpolation intervals are suggested by the $\Delta t^h(\cdot)$ in (3.1) and (3.4). We will use a Markovian interpolation, called $\psi^h(\cdot)$. Let $\{\Delta\tau_n^h, n < \infty\}$ be conditionally mutually independent and "exponential" random variables in that

$$P_{x,n}^{h,\alpha}\left(\Delta\tau_n^h \geq t\right) = e^{-t/\Delta t^h(x,\alpha)}.$$

Note that $\Delta\tau_n^h = 0$ if ξ_n^h is on the reflecting boundary ∂G_h^+. Define $\tau_0^h = 0$, and for $n > 0$, set $\tau_n^h = \sum_{i=0}^{n-1} \Delta\tau_i^h$. The τ_n^h will be the jump times of $\psi^h(\cdot)$. Now define $\psi^h(\cdot)$ and the interpolated reflection processes by

$$\psi^h(t) = x(0) + \sum_{\tau_{i+1}^h \leq t} [\xi_{i+1}^h - \xi_i^h],$$

$$Z^h(t) = \sum_{\tau_{i+1}^h \leq t} [\xi_{i+1}^h - \xi_i^h] I_{\{\xi_i^h \in \partial G_h^+\}},$$

$$z^h(t) = \sum_{\tau_{i+1}^h \leq t} E_i^h[\xi_{i+1}^h - \xi_i^h] I_{\{\xi_i^h \in \partial G_h^+\}}.$$

Define the continuous time interpolations $u_i^h(\cdot)$ of the controls analogously. Let $r_i^h(\cdot)$ denote the relaxed control representation of $u_i^h(\cdot)$. The process $\psi^h(\cdot)$ is a continuous time Markov chain. When the state is x and control pair is α, the jump rate out of $x \in G_h$ is $1/\Delta t^h(x,\alpha)$. So the conditional mean interpolation interval is $\Delta t^h(x,\alpha)$; i.e., $E_{x,n}^{h,\alpha}[\tau_{n+1}^h - \tau_n^h] = \Delta t^h(x,\alpha)$.

Define $\tilde{z}^h(\cdot)$ by $Z^h(t) = z^h(t) + \tilde{z}^h(t)$. This representation splits the effects of the reflection into two parts. The first is composed of the "conditional mean" parts $E_i^h[\xi_{i+1}^h - \xi_i^h] I_{\{\xi_i^h \in \partial G_h^+\}}$, and the second is composed of the perturbations about these conditional means [22, Section 5.7.9]. Both components can change only at t where $\psi^h(t)$ can leave G_h. Suppose that at some time t, $Z^h(t) - Z^h(t-) \neq 0$, with $\psi^h(t-) = x \in G_h$. Then by (3.4), $z^h(t) - z^h(t-)$ points in a direction in $d(N_h(x))$ where $N_h(x)$ is a neighborhood with radius that goes to zero as $h \to 0$. The process $\tilde{z}^h(\cdot)$ is the "error" due to the centering of the increments of the reflection term about their conditional means and has bounded (uniformly in x, h) second moments and it converges to zero, as will be seen in Theorem 3.1. By (A2.1), (A2.2), and the local consistency condition (3.4), we can write (modulo an asymptotically negligible term)

$$z^h(t) = \sum_i d_i y_i^h(t),$$

15

where $y_i^h(0) = 0$, and $y_i^h(\cdot)$ is nondecreasing and can increase only when $\psi^h(t)$ is arbitrarily close (as $h \to 0$) to the ith face of ∂G.

A representation for $\psi^h(\cdot)$. The process $\psi^h(\cdot)$ has a representation which resembles (2.4), and is useful in the convergence proofs. Let $\xi_0^h = x$. By [22, Sections 5.7.3 and 10.4.1], we can write

$$
\psi^h(t) = x + \int_0^t b(\psi^h(s), u^h(s))ds \\
+ \int_0^t \sigma(\psi^h(s))dw^h(s) + Z^h(s) + \epsilon^h(s),
\tag{3.5}
$$

where $\psi^h(t) \in G$. The process $\epsilon^h(\cdot)$ is due to the $o(\cdot)$ terms in (3.1) and is asymptotically unimportant in that, for any T, $\lim_h \sup_{x,u^h} \sup_{s \leq T} E_x^{h,u^h} |\epsilon^h(s)|^2 = 0$. The process $w^h(\cdot)$ is a martingale with respect to the filtration induced by $(\psi^h(\cdot), u^h(\cdot), w^h(\cdot))$, and converges weakly to a standard (vector-valued) Wiener process. The $w^h(t)$ is obtained from $\{\psi^h(s), s \leq t\}$. All of the processes in (3.5) are constant on the intervals $[\tau_n^h, \tau_{n+1}^h)$.

Let $|z^h|(T)$ denote the variation of the process $z^h(\cdot)$ on the time interval $[0, T]$. Then we have the following theorem from [22].

Theorem 3.1. (Theorem 11.1.3 and (5.7.5))[22].) *Assume* (A2.1), (A2.2), *the local consistency conditions, and let* $b(\cdot)$ *and* $\sigma(\cdot)$ *be bounded and measurable. Then for any* $T < \infty$, *there are* $K_2 < \infty$ *and* δ_h, *where* $\delta_h \to 0$ *as* $h \to 0$, *and which do not depend on the controls or initial condition, such that*

$$
E \, \bar{z}^{h^{-2}}(T) \leq K_2,
\tag{3.6}
$$

$$
E \sup_{s \leq T} \bar{\tilde{z}}^h(s)^{-2} = \delta_h E \, \bar{z}^{h^-}(T).
\tag{3.7}
$$

Owing to the fact that the reflection directions at any corner or edge are linearly independent, the inequalities hold for $y^h(\cdot)$ *replacing* $z^h(\cdot)$.

The cost function and upper and lower values for the discrete game. Relaxed feedback controls, when applied to the Markov chain, are equivalent to randomized controls. Let $u^h(\cdot) = (u_1^h(\cdot), u_2^h(\cdot))$ be feedback controls for the approximating chain. Then the cost is

$$
\gamma_T^h(x, u^h) = \gamma_T^h(x, u_1^h, u_2^h) = \frac{1}{T} E_x^{h,u^h} \int_0^T k_{u^h}(\psi^h(s))ds + E_x^{h,u^h} \frac{c'y^h(T)}{T}, \\
\gamma^h(u^h) = \lim_T \gamma_T^h(x, u^h).
\tag{3.8}
$$

Now suppose that $m^h(\cdot)$ represents a randomized control (as discussed above

16

(3.2)). Then the cost function can be written as

$$\gamma_T^h(x, m^h) = \gamma_T^h(x, m_1^h, m_2^h) = \frac{1}{T} E_x^{h, m^h} \int_0^T k_{m^h}(\psi^h(s)) ds + E_x^{h, m^h} \frac{c' y^h(T)}{T},$$
$$\gamma^h(m^h) = \lim_T \gamma_T^h(x, m^h).$$

(3.9)

With the relaxed feedback control representation of an ordinary feedback control, (3.8) is a special case of (3.9). Also, we can always take the controls in (3.9) to be randomized feedback.

Suppose that player 1 chooses its control first and uses the relaxed feedback (or randomized feedback) control $m_1^h(\cdot)$. Then player 2 has a maximization problem for a finite state Markov chain. The approximating chain is ergodic for any feedback control, whether randomized or not. Then, since the transition probabilities and cost rates are continuous in the control of the second player, the optimal control of the second player exists and is a pure feedback control (not randomized) [8, volume 2], [25]. The cost does not depend on the initial condition. The analogous situation holds if player 2 chooses its control first. These facts will be used in the next theorem. We use $m_i^h(\cdot)$ to denote either a randomized feedback, relaxed feedback, or the relaxed feedback representation of an ordinary feedback control. Define the upper and lower values, resp.:

$$\bar{\gamma}^{+,h} = \inf_{m_1^h} \sup_{m_2^h} \gamma^h(m_1^h, m_2^h),$$

$$\bar{\gamma}^{-,h} = \sup_{m_2^h} \inf_{m_1^h} \gamma^h(m_1^h, m_2^h).$$

Under our hypotheses, the upper and lower values might be different, although Theorem 3.2 says that they converge to the same value asymptotically. If the dynamics are separated in the sense that $P^h(x, y|\alpha)$ can be written as a function of (x, y, α_1) plus a function of (x, y, α_2), then $\bar{\gamma}^{+,h} = \bar{\gamma}^{-,h}$. [The proof is similar to that giving the analogous result in Section 4, except that the state space is discrete here.] One can choose the transition probability so that it is separated, if desired.

3.2 Convergence of the Numerical Procedure

Theorem 3.2. *Assume* (A2.1)–(A2.5) *and suppose that*[1]

$$\bar{\gamma}^+ = \bar{\gamma}^- = \bar{\gamma}.$$

(3.10)

Then

$$\bar{\gamma}^- \leq \liminf_h \bar{\gamma}^{-,h} \leq \limsup_h \bar{\gamma}^{+,h} \leq \bar{\gamma}^+.$$

(3.11)

Hence

$$\lim_h \bar{\gamma}^{+,h} = \lim_h \bar{\gamma}^{-,h} = \bar{\gamma}$$

(3.12)

[1]Equation (3.10) will be proved in the next section

and both the upper and lower values for the numerical approximation converge to the value for the original game.

Proof. Let player 1 choose its control first and let $\epsilon > 0$. Let $\overline{m}_{\epsilon,1}^+(\cdot)$ be an ϵ-smoothing of the optimal control $\overline{m}_1^+(\cdot)$ for player 1, when it chooses first, as discussed at the end of Section 2. That discussion implies that, given $\delta > 0$, there is $\epsilon > 0$ such that $\overline{m}_{1,\epsilon}^+(\cdot)$ is δ-optimal for player 1 for the original problem. Now, let player 1 use $\overline{m}_{1,\epsilon}^+(\cdot)$ on the approximating chain, either as a randomized feedback or a relaxed feedback control. Given that player 1 chooses first and uses $\overline{m}_{1,\epsilon}^+(\cdot)$, we have a simple control problem for player 2. As noted above, the optimal control for player 2 exists and is pure feedback, and we denote it by $\tilde{u}_2^h(\cdot)$, with relaxed feedback control representation $\tilde{m}_2^h(\cdot)$.

By the definition of the upper value,

$$\bar{\gamma}^{+,h} \leq \sup_{u_2^h} \gamma^h(\overline{m}_{1,\epsilon}^+, u_2^h) = \sup_{m_2^h} \gamma^h(\overline{m}_{1,\epsilon}^+, m_2^h) = \gamma^h(\overline{m}_{1,\epsilon}^+, \tilde{u}_2^h), \qquad (3.13)$$

where $u_2^h(\cdot)$ denotes an arbitrary ordinary feedback control, and $m_2^h(\cdot)$ an arbitrary randomized feedback control. The maximum value $\gamma^h(\overline{m}_{1,\epsilon}^+, \tilde{u}_2^h)$ of the control problem for player 2 with player 1's control fixed at $\overline{m}_{1,\epsilon}^+(\cdot)$ does not depend on the initial condition. Hence, without loss of generality, the corresponding continuous time interpolation $\psi^h(\cdot)$ can be considered to be stationary. Then, using the continuity in (x, α_2) of $\int_{U_1} b(x, \alpha)\overline{m}_{1,\epsilon}^+(x, d\alpha_1)$ and of $\int_{U_1} k(x, \alpha)\overline{m}_{1,\epsilon}^+(x, d\alpha_1)$ (and replacing the minimization problem by a maximization problem), yields [22, Theorem 3.1, Chapter 11] that there is a relaxed control $\tilde{r}_2(\cdot)$ for the original problem such that:[2]

$$\limsup_h \bar{\gamma}^{+,h} \leq \limsup_h \gamma^h(\overline{m}_{1,\epsilon}^+, \tilde{u}_2^h) = \gamma(\overline{m}_{1,\epsilon}^+, \tilde{r}_2) \leq \bar{\gamma}^+ + \delta. \qquad (3.14)$$

The last inequality of (3.14) follows from Theorem 2.7 and the δ-optimality of $\overline{m}_{1,\epsilon}^+(\cdot)$ in the class of relaxed feedback controls for player 1 if it chooses first.

Now, let player 2 choose first, Then there is an analogous result with analogous notation: In particular, given $\delta > 0$, there is an $\epsilon > 0$ and an ϵ−smoothing $\overline{m}_{2,\epsilon}^-(\cdot)$ of the optimal control, and a relaxed control $\tilde{r}_1(\cdot)$ for the original problem (2.4) such that

$$\liminf_h \bar{\gamma}^{-,h} \geq \liminf_h \gamma^h(\tilde{u}_1^h, \overline{m}_{2,\epsilon}^-) = \gamma(\tilde{r}_2, \overline{m}_{2,\epsilon}^-) \geq \bar{\gamma}^- - \delta. \; . \qquad (3.15)$$

Hence, since δ is arbitrary, (3.11) holds. This, with (3.10), yields the theorem. ∎

---

[2]In [22, Theorem 3.1, Chapter 11], the symbol $\mathsf{m}(\cdot)$ is used for a relaxed control and not a relaxed feedback control. That reference does not use relaxed feedback controls.

4   Existence of the Value of the Game

An approach to the proof. The existence of the value, namely (3.10), will be proved in this section. Before proceeding with the proof, we will motivate what will be needed by outlining a tentative approach. The outline is purely formal. But, later, it will be seen that the method can be carried out.

Suppose for the moment that the game for the numerical approximation has a value in that $\bar{\gamma}^{+,h} = \bar{\gamma}^{-,h}$, and let there be controls controls $\overline{m}_1^h(\cdot), \overline{m}_2^h(\cdot)$ for the numerical method (written in relaxed feedback form) which attain the value, no matter who chooses first. I.e., $\overline{m}_i^h(\cdot)$ is optimal for player i whether it chooses its control first or last. Thus,

$$\bar{\gamma}^{+,h} = \bar{\gamma}^{-,h} = \bar{\gamma}^h = \gamma^h(\overline{m}_1^h, \overline{m}_2^h). \tag{4.1}$$

Suppose also that there are relaxed feedback controls $\tilde{m}_i(\cdot)$ such that, for some subsequence of $h \to 0$,

$$\overline{m}_1^h(x, d\alpha_1)\overline{m}_2^h(x, d\alpha_2)dx \Rightarrow \tilde{m}_1(x, d\alpha_1)\tilde{m}_2(x, d\alpha_2)dx. \tag{4.2}$$

Finally, suppose that for any sequence (indexed by $h \to 0$) of relaxed feedback controls $\{m_i^h(\cdot)\}, i = 1, 2$, for which $m_1^h(x, d\alpha_1)m_2^h(x, d\alpha_2)dx$ converges weakly to, say, $m_1(x, d\alpha_1)m_2(x, d\alpha_2)dx$, we have the convergence of the costs

$$\gamma^h(m_1^h, m_2^h) \to \gamma(m_1, m_2). \tag{4.3}$$

Then by (3.11) it follows that

$$\bar{\gamma}^- \le \gamma(\tilde{m}_1, \tilde{m}_2) \le \bar{\gamma}^+.$$

We claim that, under the above hypotheses, the limit control $\tilde{m}_i(\cdot)$ is optimal for player i if it chooses first. To prove this claim one can proceed as follows. Suppose that $\tilde{m}_1(\cdot)$ is not optimal for player 1 if it chooses first, in that $\sup_{m_2} \gamma(\tilde{m}_1, m_2) > \bar{\gamma}^+$. Then there are $\delta > 0$ and $\hat{m}_2(\cdot)$ such that $\gamma(\tilde{m}_1, \hat{m}_2) \ge \bar{\gamma}^+ + 2\delta$. Following the approach in Theorem 3.2, for $\epsilon > 0$ let $\hat{m}_{2,\epsilon}(\cdot)$ be an ϵ-smoothing of $\hat{m}_2(\cdot)$. Then, for small $\epsilon > 0$, $\gamma(\tilde{m}_1, \hat{m}_{2,\epsilon}) \ge \bar{\gamma}^+ + \delta$. Then apply $\hat{m}_{2,\epsilon}(\cdot)$ to the approximating controlled process $\psi^h(\cdot)$ to get a contradiction to the optimality of $(\overline{m}_1^h(\cdot), \overline{m}_2^h(\cdot))$ for small h. Such a contradiction implies that $\sup_{m_2} \gamma(\tilde{m}_1, m_2) \le \bar{\gamma}^+$. But, the strict inequality $<$ is impossible due to the definition of the upper value. Hence $\sup_{m_2} \gamma(\tilde{m}_1, m_2) = \bar{\gamma}^+$, as desired.

To get the desired contradiction to the optimality of $(\overline{m}_1^h(\cdot), \overline{m}_2^h(\cdot))$ for small h, let h index a weakly convergence subsequence of the measures defined in the left side of (4.2). The limit must be of the form on the right side of (4.2) for some $\tilde{m}_i(\cdot), i = 1, 2$, where $\overline{m}_i^h(x, \cdot) \Rightarrow \tilde{m}_i(x, \cdot)$ for almost all $x \in G, i = 1, 2$. Apply the control pair $(\overline{m}_1^h(\cdot), \hat{m}_{2,\epsilon}(\cdot))$ to $\psi^h(\cdot)$. Then (along the chosen subsequence of h)

$$\overline{m}_1^h(x, d\alpha_1)\hat{m}_{2,\epsilon}(x, d\alpha_2)dx \Rightarrow \tilde{m}_1(x, d\alpha_1)\hat{m}_{2,\epsilon}(x, d\alpha_2)dx.$$

Since (4.3) implies that $\gamma^h(\overline{m}_1^h, \hat{m}_{2,\epsilon}) \to \gamma(\tilde{m}_1, \hat{m}_{2,\epsilon})$, for small enough ϵ and h, we must have $\gamma^h(\overline{m}_1^h, \hat{m}_{2,\epsilon}) \geq \bar{\gamma}^{+,h} + \delta/2$, which is a contradiction to the optimality of $\overline{m}_1^h(\cdot)$. We can now conclude that

$$\sup_{m_2} \gamma(\tilde{m}_1, m_2) = \bar{\gamma}^+ = \gamma(\tilde{m}_1, \tilde{m}_2). \tag{4.4}$$

Thus, if player 1 chooses its control first and uses its optimal control $\tilde{m}_1(\cdot)$, then $\tilde{m}_2(\cdot)$ is optimal for player 2. By repeating the procedure with the order of the players reversed, we can finally conclude that, if (4.1)–(4.3) hold (at least for some subsequence of h), then (3.10) holds.

The approach outlined above for proving (3.10) is attractive. But it *cannot* work for the class of processes $\psi^h(\cdot)$ which are used for the actual Markov chain approximation numerical method in Section 3, since for each h, the state space is only some finite set. Hence, the controls are not defined for all $x \in G$, and the transition function is not mutually absolutely continuous with respect to Lebesgue measure. However, in this section we are concerned only with proving (3.10), and not with the numerical procedure. Thus, we can use the approach which was outlined above for an appropriately chosen alternative approximating process for which (3.11) also holds. A discrete time process will be constructed for which (3.11) and (4.1)–(4.3) hold. This process is to be used solely to prove (3.10). It is not suitable for numerical solution. For future use, note that if the $\overline{m}_i^h(\cdot), i = 1, 2$, are relaxed feedback controls for each h and the $\overline{m}_i^h(x, \cdot)$ are defined for almost all x, then there is always a subsequence and relaxed feedback controls $\tilde{m}_i(\cdot), i = 1, 2$, for which (4.2) holds.

An alternative approximating process

An alternative approximating process. To get the approximating process, time will be discretized but not space. Let $\Delta > 0$ denote the time discretization interval. We need to construct process whose n-step transition functions $P^\Delta(x, n\Delta, \cdot | \alpha)$ have densities that are mutually absolutely continuous with respect to Lebesgue measure, uniformly in (Δ, control, $t_0 \leq n\Delta \leq t_1$) for any $0 < t_0 < t_1 < \infty$.

Consider the following procedure. Start with the process (2.4), but with the controls held constant on the intervals $[l\Delta, l\Delta + \Delta), l = 0, 1, \ldots$. The discrete approximation will be the samples at times $l\Delta, l = 0, 1, \ldots$. The controls are chosen at $t = 0$, with one of the players selected to choose first, just as for the original game. Let $u_i^\Delta(\cdot), i = 1, 2$, denote the controls, if in pure feedback (not relaxed or randomized) form. In relaxed control notation write the controls as $m_i^\Delta(\cdot), i = 1, 2$. These controls are used henceforth, whenever control is applied. The chosen controls are applied at random as follows. At each time, only one of the players will use its control. At each time $l\Delta, l = 0, 1, \ldots$, flip a fair coin. With probability 1/2, player 1 will use its control during the interval $[l\Delta, l\Delta + \Delta)$ and player 2 not. Otherwise, player 2 will use its control, and player 1 not. The values of the controls during the interval will depend on the state at its start. The optimal controls will be feedback. Define $x^\Delta(t) = x(l\Delta)$ on $[l\Delta, l\Delta + \Delta)$. For pure (not randomized or relaxed) feedback controls $u_i^\Delta(\cdot), i = 1, 2$, the system

is

$$dx = b^\Delta(x, u^\Delta(x^\Delta))dt + \sigma(x)dw + dz, \qquad (4.5a)$$

where the value of $b^\Delta(\cdot)$ is determined by the coin tossing randomization procedure at the times $l\Delta, l = 0, 1 \ldots$, In particular, at $t \in [l\Delta, l\Delta+\Delta)$, $b^\Delta(x, m^\Delta(x^\Delta))$ is $2b_i(x(t), u_i^\Delta(x^\Delta(t)))$, for either $i = 1$ or $i = 2$ according to the random choice made at $l\Delta$. If the control is relaxed feedback, then write the model as

$$dx = b^\Delta(x, m^\Delta(x^\Delta))dt + \sigma(x)dw + dz, \qquad (4.5b)$$

where at $t \in [l\Delta, l\Delta+\Delta)$, $b^\Delta(x, m^\Delta(x^\Delta))$ is $2\int_{U_i} b_i(x(t), \alpha_i)m_i^\Delta(x(l\Delta), d\alpha_i)$, for either $i = 1$ or $i = 2$ according to the random choice made at $l\Delta$. Following the Girsanov transformation based usage in (2.4), the Wiener process $w(\cdot)$ should be indexed by the controls $u^\Delta(\cdot)$ or $m^\Delta(\cdot)$, but we omit it for notational simplicity.

Let $E_{x(l\Delta)}^{\Delta,i,\alpha_i}$ denote the expectation of functionals on $[l\Delta, l\Delta+\Delta)$ when player i acts on that interval and uses control action α_i. Let $P_i^\Delta(x, \cdot|\alpha_i)$ denote the the measure of $x(\Delta)$, given that the initial condition is x, player i acts and uses control action α_i. The conditional mean increment in the total cost function on the time interval $[l\Delta, l\Delta + \Delta)$ is, for $u_i^\Delta(x(l\Delta)) = \alpha_i, i = 1, 2$,

$$C^\Delta(x(l\Delta), \alpha) = \frac{1}{2} \sum_{i=1,2} E_{x(l\Delta)}^{\Delta,i,\alpha_i} \left[\int_{l\Delta}^{l\Delta+\Delta} 2k_i(x(s), \alpha_i))ds + c'\left(y(l\Delta + \Delta) - y(l\Delta)\right) \right].$$

$$(4.6)$$

Note that $C^\Delta(x, \alpha)$ is the sum of two terms, one depending on (x, α_1) and the other on (x, α_2). The weak sense uniqueness of the solution to (2.4) for any control and initial condition implies the following result.

Theorem 4.1. *Assume (A2.1)–(A2.5). Then for each $\Delta > 0$, $C^\Delta(\cdot)$ is continuous and the measures $P_i^\Delta(\cdot)$ are weakly continuous in that for any bounded and continuous real-valued function $f(\cdot)$, $\int f(y)P_i^\Delta(x, dy|\alpha)$ and $C^\Delta(x, \alpha)$ are continuous in (x, α).*

The reason for choosing the acting controls at random at each time $l\Delta, l = 0, 1, \ldots$, is that the randomization "separates" the cost rates and dynamics in the controls for the two players. By separation, we mean that both the cost function and transition function are the sum of two terms, one depending on (x, α_1) and the other on (x, α_2). This separation is important since it gives the "Isaacs condition " which is needed to assure the existence of a value for the game for the discrete time process, as seen in Theorem 4.2. Proceeding formally at this point, let $\mu_{m^\Delta}^\Delta(\cdot)$ denote the invariant measure under the control $m^\Delta(\cdot)$. Define the stationary cost increment

$$\lambda^\Delta(m^\Delta) = \int_G \mu_{m^\Delta}^\Delta(dx) \int_U C(x, \alpha)m^\Delta(x, d\alpha) .$$

Note that, due to the scaling, $\lambda^\Delta(m^\Delta)$ is an average over an interval of length Δ: hence $\lambda^\Delta(m^\Delta) = \Delta\gamma^\Delta(m^\Delta)$. Suppose for the moment that there is an

optimal control $\overline{m}_i^\Delta(\cdot), i = 1, 2$, for each $\Delta > 0$ and define $\overline{\lambda}^\Delta = \lambda^\Delta(\overline{m}^\Delta)$. The "separation" is easily seen from the formal Isaacs equation for the value of the discrete time problem, namely,

$$\bar\lambda^\Delta + \overline{g}^\Delta(x) =$$
$$\inf_{\alpha_1} \sup_{\alpha_2} \left[\frac{1}{2} \int \overline{g}^\Delta(x + y) P_1^\Delta(x, dy | \alpha_1) + \frac{1}{2} \int \overline{g}^\Delta(x + y) P_2^\Delta(x, dy | \alpha_2) + C^\Delta(x, \alpha) \right], \tag{4.7}$$

where $\overline{g}^\Delta(\cdot)$ is the relative value or potential function.

Theorem 4.2. *Assume* (A2.1)–(A2.5). *Then* (3.10) *holds.*

Proof. We will work with the approximating process $x(l\Delta), l = 0, 1, \ldots$ just described, where $x(\cdot)$ is defined by (4.5) with the piecewise constant control, and verify the conditions imposed in the formal discussion at the beginning of the section. Results from [21] will be exploited whenever possible. The result (3.11) holds (with Δ replacing h) for the same reasons that it holds for the numerical approximating process of the last section. For any sequence of relaxed controls $m_i^\Delta(\cdot), i = 1, 2$, there is a subsequence (indexed by Δ) and $\tilde{m}_i^\Delta(\cdot), i = 1, 2$, such that

$$m_1^\Delta(x, d\alpha_1) m_2^\Delta(x, d\alpha_2) dx \Rightarrow \tilde{m}_1(x, d\alpha_1) \tilde{m}_2(x, d\alpha_2) dx.$$

One needs to show the analog of (4.3), namely (along the same subsequence, indexed by Δ)

$$\gamma^\Delta(m^\Delta) \to \gamma(\tilde{m}). \tag{4.8}$$

The process $\{x(l\Delta)\}$ based on (4.5) inherits the crucial properties of (2.4), as developed in [21, Chapter 4] and summarized in Subsection 2.2. In particular, for each positive Δ and n the n−step transition probability $P^\Delta(x, n\Delta, \cdot | m^\Delta)$ is mutually absolutely continuous with respect to Lebesgue measure, uniformly in the control and in $x \in G$, $n\Delta \in [t_0, t_1]$, for any $0 < t_0 < t_1 < \infty$, and it is a strong Feller process. The invariant measures are mutually absolutely continuous with respect to Lebesgue measure, again uniformly in the control. Then the proof of (4.8) is very similar to the corresponding proof for (2.4) given in [21, Theorem 4.3, Chapter 4] and the details are omitted. There are controls $\overline{m}_1^{\Delta,+}(\cdot)$ which are optimal if player 1 chooses its control first (i.e., for the upper value), and $\overline{m}_2^{\Delta,-}(\cdot)$ which are optimal if player 2 chooses its control first (i.e., for the lower value).

We will concentrate on showing the analog of (4.1), namely,

$$\bar\gamma^{+,\Delta} = \bar\gamma^{-,\Delta}. \tag{4.9}$$

By the (uniform in the controls) mutual absolute continuity of the one step transition probabilities for each $\Delta > 0$, the process satisfies a Doeblin condition, uniformly in the control. Hence it is uniformly ergodic, uniformly in the control) [23, Theorems 16.2.1 and 16.2.3]. In particular it follows that there are constants

K_Δ and ρ_Δ, with $\rho_\Delta < 1$ such that

$$\sup_{x,m^\Delta} \left| E_x^{\Delta,m^\Delta} \int_U C(x(n\Delta),\alpha)m^\Delta(x(n\Delta),d\alpha) - \lambda^\Delta(m^\Delta) \right| \le K_\Delta \left[\rho_\Delta\right]^n,$$

where $\lambda^\Delta(m^\Delta)$ is defined above (4.7).

Define the relative value function

$$g^\Delta(x,m^\Delta) = \sum_{l=0}^{\infty} \left[E_x^{\Delta,m^\Delta} C(x(l\Delta), m^\Delta(x(n\Delta)) - \lambda^\Delta(m^\Delta) \right].$$

The summands converge to zero exponentially, uniformly in $(x,m^\Delta(\cdot))$. Also, by the strong Feller property the summands (for $l > 0$) are continuous. Define $g^{\Delta,+}(x) = g^\Delta(x,\overline{m}^{\Delta,+})$ and $g^{\Delta,-}(x) = g^\Delta(x,\overline{m}^{\Delta,-})$. Then, a direct evaluation yields

$$\bar\lambda^{\Delta,+} + g^{\Delta,+}(x) = E_x^{\Delta,\overline{m}^{\Delta,+}} \left[g^{\Delta,+}(x(\Delta)) + C^\Delta(x,\overline{m}^{\Delta,+}(x)) \right]. \tag{4.10}$$

Next we show that under $\overline{m}_1^{\Delta,+}(\cdot)$ (and for almost all x)

$$\bar\lambda^{\Delta,+} + g^{\Delta,+}(x) = \sup_{\alpha_2} \left[E_x^{\Delta,\overline{m}_1^{\Delta,+},\alpha_2} g^{\Delta,+}(x(\Delta)) + C^\Delta(x,\overline{m}_1^{\Delta,+}(x),\alpha_2) \right]. \tag{4.11}$$

By (4.10), (4.11) holds for almost all x with the equality replaced by the inequality $\le$. The function in brackets in (4.11) is continuous in α_2, uniformly in $x \in G$. Suppose that (4.11) does not hold on a set $A \subset G$ of Lebesgue measure $l(A) > 0$. Let $\tilde{m}_2^\Delta(\cdot)$ denote the (relaxed feedback control representation of the) maximizing control in (4.11). Then

$$\bar\lambda^{\Delta,+} + g^{\Delta,+}(x) \le \left[E_x^{\Delta,\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta} g^{\Delta,+}(x(\Delta)) + C^\Delta(x,\overline{m}_1^{\Delta,+}(x),\tilde{m}_2^\Delta(x)) \right],$$
$$\tag{4.12}$$

with strict inequality for $x \in A$. Now, integrate both sides of (4.12) with respect to the invariant measure $\mu^\Delta_{\{\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta\}}(\cdot)$ corresponding to the control $(\overline{m}_1^\Delta(\cdot), \tilde{m}_2^\Delta(\cdot))$ and note that

$$\int g^{\Delta,+}(x)\mu^\Delta_{\{\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta\}}(dx) = \int \left[E_x^{\Delta,\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta} g^{\Delta,+}(x(\Delta)) \right] \mu^\Delta_{\{\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta\}}(dx).$$
$$\tag{4.13}$$

Also, by definition,

$$\lambda^\Delta(\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta) = \int C^\Delta(x,\overline{m}_1^{\Delta,+}(x),\tilde{m}_2^\Delta(x))\mu^\Delta_{\{\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta\}}(dx).$$

Then, canceling the terms in (4.13) from the integrated inequality and using the fact that the invariant measure is mutually absolutely continuous with respect to Lebesgue measure yields $\bar\lambda^{\Delta,+} < \lambda^\Delta(\overline{m}_1^{\Delta,+},\tilde{m}_2^\Delta)$, which contradicts the optimality of $\overline{m}_2^{\Delta,+}(\cdot)$ for player 2, if player 1 selects its control first. Thus, (4.11) holds.

23

Next, given that (4.11) holds, let us show that for almost all x

$$\bar{\lambda}^{\Delta,+} + g^{\Delta,+}(x) = \inf_{\alpha_1} \sup_{\alpha_2} E_x^{\Delta,\alpha_1,\alpha_2} \left[g^{\Delta,+}(x(\Delta)) + C^\Delta(x,\alpha_1,\alpha_2) \right], \qquad (4.14)$$

By (4.11), this last equation holds if $\overline{m}_1^{\Delta,+}(\cdot)$ replaces α_1 and the inf is dropped. Suppose that (4.14) is false. Then there are $A \in G$ with $l(A) > 0$ and $\epsilon > 0$ such that for $x \in A$ the equality is replaced by the inequality $\geq$ plus ϵ, with the inequality $\geq$ holding for almost all other $x \in G$. More particularly, let $\hat{m}_1^{\Delta,+}(\cdot)$ denote the minimizing control for player 1 in (4.14). Then we have, for almost all x and any $m_2^\Delta(\cdot)$,

$$\bar{\lambda}^{\Delta,+} + g^{\Delta,+}(x) \geq E_x^{\Delta,\hat{m}_1^\Delta,m_2^\Delta} \left[g^{\Delta,+}(x(\Delta)) + C^\Delta(x,\hat{m}_1^\Delta(x),m_2^\Delta(x)) \right] + \epsilon I_{\{x \in A\}}, \tag{4.15}$$

Now, repeating the procedure used to prove (4.11), integrate both sides of (4.15) with respect to the invariant measure associated with $(\hat{m}_1^\Delta(\cdot), m_2^\Delta(\cdot))$, use the fact that the invariant measure is mutually absolutely continuous with respect to Lebesgue measure, u niformly in the controls, and cancel the terms which are analogous to those in (4.13), to get that

$$\bar{\lambda}^{\Delta,+} > \sup_{m_2^\Delta} \lambda^\Delta(\hat{m}_1^\Delta, m_2^\Delta).$$

This implies that $\overline{m}_1^{\Delta,+}(\cdot)$ is not optimal for player 1 if it selects its control first, a contradiction. Thus, (4.14) holds. The analogous procedure can be carried out for the lower value where player 2 selects its control first..

Now the fact that the dynamics and cost rate are separated in the control implies that $\inf_{\alpha_1} \sup_{\alpha_2} = \sup_{\alpha_2} \inf_{\alpha_1}$ in (4.14). Thus, (4.14) holds with the order of the sup and inf inverted. By working with the equation (4.14) with the sup and inf inverted and following an argument similar to that used to prove (4.14), one can show that $\bar{\lambda}^{\Delta,+} = \bar{\lambda}^{\Delta,-}$ and that $\overline{m}_i^\Delta(\cdot)$ is optimal for player i whether it selects first or last. The rest of the details are left to the reader. ∎

References

[1] E. Altman and H.J. Kushner. Admission control for combined guaranteed performance and best effort communications systems under heavy traffic. *SIAM J. Control and Optimiz.*, 37:1780–1807, 1999.

[2] J. A. Ball, M. Day, and P. Kachroo. Robust feedback control of a single server queueing system. *Math. of Control, Signals and Systems*, 12:307–345, 1999.

[3] J. A. Ball, M. Day, P. Kachroo, and T. Yu. Robust L_2−gain for nonlinear systems with projection dynamics and and input constraints: an example from traffic control. *Automatica*, 35:429–444, 1999.

[4] M. Bardi, S. Bottacin, and M. Falcone. Convergence of discrete schemes for discontinuous value functions of pursuit-evasion games. In G.I. Oldser, editor, *New Trends in Dynamic Games and Applications*. Birkhäuser, Boston, 1995.

[5] M. Bardi, M. Falcone, and P. Soravia. Fully discrete schemes for the value function of pursuit-evasion games. In T. Basar and A. Haurie, editors, *Advances in Stochastic Games and Applications*. Birkhäuser, Boston, 1994.

[6] M. Bardi, M. Falcone, and P. Soravia. Numerical methods for pursuit-evasion games via viscosity solutions. In M. Bardi, T.E.S. Raghavan, and T. Parthasarathy, editors, *Stochastic and Differential Games: Theory and Numerical Methods*. Birkhäuser, Boston, 1998.

[7] T. Basar and P. Bernhard. H_∞-*Optimal Control and Related Minimax problems*. Birkhäuser, Boston, 1991.

[8] D. M. Bertsekas. *Dynamic Programming and Optimnal Control*. Athena-Scientific, Belmont, Mass., 1995.

[9] P. Billingsley. *Convergence of Probability Measures; Second edition*. Wiley, New York, 1999.

[10] V.S. Borkar. *Optimal Control of Diffusion Processes*. Longman Scientific and Technical, Harlow, Essex, UK, 1989.

[11] N. Dunford and J.T. Schwartz. *Linear Operators, Part 1: General Theory*. Wiley-Interscience, New York, 1966.

[12] P. Dupuis and H. Ishii. On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications. *Stochastics and Stochastic Rep.*, 35:31–62, 1991.

[13] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[14] W.F. Fleming. Generalized solutions in optimal stochastic control. In P.T. Liu, E. Roxin, and R. Sternberg, editors, *Differential Games and Control Theory: III*, pages 147–165. Marcel Dekker, 1977.

[15] W.H. Fleming and W. McEneaney. Risk-sensitive control on an infinie time horizon. *SIAM J. on Control and Optimiz.*, 33:1881–1815, 1995.

[16] J.M. Harrison and R.J. Williams. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics and Stochastics Rep.*, 22:77–115, 1987.

[17] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 1988.

[18] H. J. Kushner. Numerical methods for stochastic differential games. Brown University, Applied Math., Report; to appear in SIAM J. on Control and Optim., 2001.

[19] H.J. Kushner. *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*. Academic Press, New York, 1977.

[20] H.J. Kushner. Numerical methods for stochastic control problems in continuous time. *SIAM J. Control Optim.*, 28:999–1048, 1990.

[21] H.J. Kushner. *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*. Springer-Verlag, Berlin and New York, 2001.

[22] H.J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, Berlin and New York, 1992. Second edition, 2001.

[23] S.P. Meyn and R.I. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, Berlin and New York, 1994.

[24] O. Pourtallier and M. Tidball. Approximation of the value function for a class of differential games with target. Research report 2942, INRIA, France, 1996.

[25] M.L. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.

[26] M.I. Reiman and R.J. Williams. A boundary property of semimartingale reflecting Brownian motions. *Prob. Theory Rel. Fields*, 77:87–97, 1988.

[27] M.R. Reiman. Open queueing networks in heavy traffic. *Math. Oper. Res.*, 9:441–458, 1984.

[28] M. Tidball. Undiscounted zero-sum differential games with stopping times. In G.J. Oldser, editor, *New Trends in Dynamic Games and Applications*. Birkhäuser, Boston, 1995.

[29] M. Tidball and R.L.V. González. Zero-sum differential games with stopping times: Some results and about its numerical resolution. In T. Basar and A. Haurie, editors, *Advances in Dynamic Games and Applications*. Birkhäuser, Boston, 1994.